Leveraging study of robustness and portability of spoken language understanding systems across languages and domains: the PORTMEDIA corpora

Fabrice Lefèvre, Djamel Mostefa, Laurent Besacier, Yannick Estève, Matthieu Quignard, Nathalie Camelin, Benoit Favre, Bassam Jabaian, Lina Maria Rojas Barahona

Leveraging study of robustness and portability of spoken language understanding systems across languages and domains: the PORTMEDIA corpora

Fabrice Lefèvre[⋆], Djamel Mostefa[†], Laurent Besacier[∗], Yannick Estève[‡], Matthieu Quignard[◦], Nathalie Camelin[‡], Benoit Favre[††], Bassam Jabaian[⋆∗], Lina Rojas-Barahona[◦]

[⋆]University of Avignon, LIA-CERI, France, {fabrice.lefevre,bassam.jabaian}@univ-avignon.fr
[†]Evaluation and Language resources Distribution Agency, France, mostefa@elda.org
[∗] LIG, Grenoble, France, laurent.besacier@imag.fr
[‡] LIUM, Le Mans, France, {yannick.esteve,nathalie.camelin}@univ-lemans.fr
[◦] LORIA, Nancy, France, {matthieu.quignard,lina.rojas}@loria.fr
[††] LIF, Marseille, France, benoit.favre@lif.univ-mrs.fr

Abstract

The PORTMEDIA project is intended to develop new corpora for the evaluation of spoken language understanding systems. The newly collected data are in the field of human-machine dialogue systems for tourist information in French in line with the MEDIA corpus. Transcriptions and semantic annotations, obtained by low-cost procedures, are provided to allow a thorough evaluation of the systems' capabilities in terms of robustness and portability across languages and domains. A new test set with some adaptation data is prepared for each case: in Italian as an example of a new language, for ticket reservation as an example of a new domain. Finally the work is complemented by the proposition of a new high level semantic annotation scheme well-suited to dialogue data.

Keywords: spoken language understanding system, human-machine dialogue corpus, portability

1. Introduction

With the ever growing spread of human-machine communications (call centers, telephone services, smartphones, etc.), Spoken Language Understanding (SLU) received an increasing interest over the past few years. SLU systems have been deployed on real world applications but with a limited success so far. Firstly SLU systems are usually available in only one language and don't support multilinguality. Secondly existing services with SLU components are very constrained and restricted by the task or the domain of application. Finally the quality of the user-system interaction is still far from being enjoyable and natural. An important aspect of the SLU system's development is the possibility to evaluate them individually of the other system's components (as in (Hahn et al., 2010)). The MEDIA evaluation campaign has allowed to setup an evaluation paradigm based on a flat and segmental conceptual annotation and to collect and process data accordingly (Bonneau-Maynard et al., 2005). The PORTMEDIA project is a follow-up of the MEDIA project (2003-2007) for which a 1258 dialogue French corpus for the *touristic domain* was produced. The PORTMEDIA project tries to address some remaining issues by developing new corpora targeting three distinct but complementary objectives:

- **Robustness of SLU systems to recognition errors.** Automatic speech recognition errors have to be included in the understanding process and thus automatic transcripts must be made available with training data to evaluate the robustness of SLU systems to recognition errors.
- **Portability across domains and languages.** SLU systems are very language and domain dependent. Adapting a system to a new domain or a new language usually requires to collect a new large speech database of the target domain/language. This process is costly *per se* and involves an important development effort. PORTMEDIA intends to leverage the evaluation of the genericity and adaptability of new approaches for SLU systems by providing data allowing contrastive experiments of portability from a source to a target language or domain.
- **High level semantic knowledge representation.** High Level Semantic (HLS) representation is needed to take into account the semantic composition process occurring inside and between consecutive user's interactions. HLS is viewed as a mean to improve the expressiveness of the annotation conveying the speaker's intentions and expectations. Starting from the flat segmental conceptual segmentation proposed in the MEDIA project, PORTMEDIA investigates the settlement of a new highly relevant hierarchical annotation scheme.

2. The PORTMEDIA spoken dialogue databases

State of the art of SLU systems use statistical models that need to be trained on large corpora of dialogues (*e.g.* acoustic and language models, conceptual models, dialogue models for some systems). There are very few publicly available corpora of Human-Computer spoken dialogues. Indeed, unlike other kind of speech, like read speech or broadcast news, the only large spoken dialogue corpus publicly available through LDC[1] are Human-Human dialogues (*e.g.*

[1]Linguistic Data Consortium

Name	Lang	Domain	#Dialogues	#Hours	#Words	#Lexicon size	#Semantic tags
MEDIA	fr	touristic information	1 258	71h	438k	3203	53 844
PM-DOM	fr	ticket reservation	700	40.5	293k	3065	17 859
PM-LANG	it	touristic information	604	50	218k	3253	20 504

Table 1: Corpus statistics for MEDIA, PM-LANG and PM-DOM

CallHome, CallFriend, Fisher corpus). The lack of data can be explained by the difficulty of collecting such corpora. Either a SLU system needs to be set-up to collect the data or one can simulate a machine through a Wizard-of-Oz protocol. Once the data is collected, annotations can be performed. While orthographic transcription is a well defined task, setting-up a semantic annotation framework is non trivial and requires a lot of expertise.

The largest Human-Computer dialogue corpora have been collected by the telecom industry on prototypes or deployed systems like the AT&T *how may I help you?* (Gorin et al., 1997) but are unfortunately not available publicly.

PORTMEDIA has produced two new corpora to study robustness and adaptability of SLU systems among languages and domains. The first corpus is made of 604 dialogues for the Italian language on the touristic domain. The second one includes 700 dialogues in French for ticket reservation in the scope of the 2010 Festival d'Avignon. Detailed statistics on the corpus are given in Table 1.

2.1. PM-LANG: the PORTMEDIA Italian corpus

The Italian database, called PM-LANG, has been recorded, transcribed and annotated with the same specifications and configurations as the MEDIA corpus. The only difference between the French MEDIA corpus and the Italian PM-LANG corpus is the language. In 2004, to collect the MEDIA corpus, 250 scenarios were used and a telephony recording platform was set-up. The recording platform included an automatic sentence generator to help the agents in their responses. For the Italian database, the same tools, protocols, scenarios and constraints were used to collect the dialogues. The only adaptation was to translate the prompts and the scenarios from French to Italian, but no changes were made to the content of the scenarios. The recording protocol is detailed in (Devillers et al., 2004). One hundred and thirty native Italian speakers were recruited and asked to call the vocal tourist telephony information server. Each speaker believes she or he is talking to a machine whereas in fact he or she is talking to a human being who simulates the behavior of a tourist information server. In order to do this simulation, the operator uses a graphical tool which generates responses that are spoken out to the caller. The generated sentences are obtained by completing a sentence template with the information obtained by consulting a real touristic information website and taking into account the caller's request. In order to have enough variety for the dialogues, speakers were asked to follow predefined scenarios for their requests. Eight levels of complexity were defined for the scenarios with different constraints (price, date, hotel standing, equipment, etc) from a simple reservation to multiple reservations and organisation of conferences. In addition to the scenarios, the operator can be cooperative or not, give explicit feedback or not and simulate some recognition errors.

Technically, the recording platform is made of a PC with a 4-lines telephony acquisition card and recording software. The dialogues are recorded in Microsoft RIFF stereo wav files with separate channels for the operator and the caller. The annotation procedure and guidelines are described the same as for MEDIA and are described in (Bonneau-Maynard et al., 2005). The resulting database is a corpus of 604 dialogues orthographically transcribed and semantically annotated.

2.2. PM-DOM: the PORTMEDIA new domain corpus

To study adaptation techniques among domains, we developed a French Human-Computer dialogue corpus (PM-DOM) with the same paradigm and specifications as MEDIA but on a different domain. While MEDIA was addressing *the touristic information domain*, the French PORTMEDIA corpus addresses *ticket reservation* within the 2010 Festival d'Avignon. We tried to remain as close as possible to the specifications, tools and paradigm of MEDIA and to minimize the differences between the two databases. The only adaptation was to create new scenarios for the callers, to adapt the dialogue management system for the agents and to develop the ontology of the domain for semantic annotation. The same recording platform and protocols were used. The dialogues are stored in 2 channels Microsoft RIFF wav files, with one channel for the caller and one channel for the speaker. This corpus is made of 700 dialogues with orthographic transcription and semantic annotation.

Transcriptions of all databases were done with Transcriber[2]. In addition to the verbatim transcriptions, markers for speaker and environment noises are added to the transcriptions. The transcriptions are segmented into speaker turns and at each speaker breath. Since the recordings were done in stereo with separate channels for the caller and the operator, it is easy to separate the speech of the caller and the operator.

The semantic annotation was done with Semantizer[3]. The semantic annotation representation for MEDIA, PM-LANG and PM-DOM uses a flat 3-tuple representation (mode, concept, normalised value) with:

- the mode: "+" (affirmative) / "-" (negative) / "?" interrogative / "~" optional"
- the name of the concept ("date", "name", "number", etc)
- the normalised value of the concept ("11/03/2012","hotel Hilton", "4")

[2]http://trans.sourceforge.net

[3]http://perso.limsi.fr/Individu/hbm/

For instance, here is a representation of a speaker's utterance taken from PM-LANG:

```
speaker:  prenotare vorrei prenotare /
un albergo / per 4 notti
+:command-tache:reservation
+:objetBD:hotel
+:sejour-nbNuit:4
```

The order of the 3-tuples in the semantic representation follows the order of the utterance. The detailed ontology for PM-LANG is described in (Bonneau-Maynard et al., 2005). For PM-DOM, we adopted a down-to-top approach for building the ontology. We started with basic concepts and asked the annotators to propose new concepts during the annotation process. Regular meetings were organised to discuss and harmonize the ontology and to revise the annotated dialogues. This iterative process allowed us to build an ontology for the new domain efficiently.

During the semantic annotation process of PM-LANG and PM-DOM, Inter-annotator Agreements (IAgs) were measured for quality assurance. For this purpose we randomly selected several sets of dialogues that were annotated twice by two annotators. The annotators were not aware that some dialogues were given to different people for measuring the IAgs.

Three IAgs were conducted for each database. The error rate was computed by aligning the sequence of 3-tuples of each annotator and counting the number of insertion, deletion, substitution in the same way as it is done in speech recognition with word error rate (Hunt, 1990). The metric was very strict since the order of the annotation is taken into account and any difference in the mode, the concept or the value for a semantic segment is counted as an error.

If we compare the annotations of the three corpora, reported in Table 2, we can observe that the most used concepts are very generic like answer (yes/no), localization, date or number.

3. Pre-transcription and semantic pre-annotation

Transcriptions and semantic annotations of PM-LANG and PM-DOM have been done in a semi-automatic way. Thanks to the availability of the MEDIA corpus, LIUM developed a speech recognizer to transcribe the PM-DOM corpus. Once the data was transcribed, it has been manually corrected by human transcribers, thanks to the tool Semantizer.

The development of the speech recognizer was done iteratively. A first set of dialogues was automatically transcribed and then corrected manually. The corrections were sent back to LIUM to retrain models and improve the speech recognition system. Then a new batch of data was automatically transcribed, manually corrected and sent back for system improvement. More details on the LIUM speech recognition system can be found in 3.1.

For semantic annotations, LIA/LIG pre-annotated both databases automatically. The pre-annotation was then corrected manually by two linguists per language. The same iteration process as for the transcription task took place for semantic annotations. Details on the LIA/LIG SLU modules are given in Section 3.2.

3.1. Speech recognition system for French

The LIUM-PM ASR system is a five-pass system based on the open-source CMU Sphinx system (version 3 and 4), similar to the LIUM'08 French ASR system described in (Deléglise et al., 2009): the first pass uses generic acoustic models with 6500 tied states and 22 Gaussians per state, and a 3-gram language model. The best hypotheses generated by the first pass are used to compute a CMLLR transformation for each speaker. Using Speaker Adaptive Training (Anastasakos et al., 1997) and Minimum Phone Error (Povey and Woodland, 2002) acoustic models (with 7500 tied states and 22 Gaussians per state) and Constrained maximum likelihood linear regression (Digalakis et al., 1995) transformations, the second pass generates word-graphs. In the third pass, word-graphs are rescored by using a better inter-word acoustic scoring: same acoustic and linguistic models as the ones used during the previous pass are used, but the decoding algorithm is different.

The fourth pass consists in recomputing with a 4-gram language model the linguistic scores of the updated word-graphs of the third pass. The last pass generates a confusion network from the word-graphs and applies the consensus method to extract the final one-best hypothesis (Mangu et al., 2000).

Acoustic models were estimated on the ESTER 1 corpus (Galliano et al., 2005), the ESTER 2 corpus (Galliano et al., 2009) and the EPAC corpus (Estève et al., 2010): this constitutes a data set of almost 280 hours of broadcast news. Language models and vocabulary were computed from the MEDIA corpus. In order to process the first data set of audio recordings of the PM-DOM data, the LIUM-PM ASR system used a vocabulary and a language model estimated on the training data of the MEDIA corpus. Vocabulary size was about 5000 words. The word error rate of this ASR system on the MEDIA test corpus was 25.2% on the caller utterances.

Table 3 is populated with the word error rates reached by the LIUM-PM ASR system for each iteration of the pre-transcription process presented in Section 3. Four data sets were automatically processed, leading to a total of 700 dialogues. For each iteration, language models and ASR vocabulary were updated according to the manual correction of the previous iteration.

3.2. SLU systems for French and Italian

In order to perform the semantic pre-annotation of the French and Italian corpora, an SLU tagger is trained based on statistical models: the conditional random fields (CRF) (Lafferty et al., 2001). A semantically annotated corpus is needed to train this tagger.

For the French PM-DOM corpus, semantic pre-annotation was obtained combining MEDIA understanding models and a new Named Entities system.

In a first step, the CRF models trained on the entire MEDIA corpus were directly applied on the PM-DOM corpus. This is an efficient way to easily find all concepts that can be shared between both corpora due to their similar lexical constituents. For instance, concepts dealing with command, location or payment take part in both MEDIA and PM-DOM applications. As an illustration, 2262 *com-*

Rank	MEDIA	PM-LANG	PM-DOM
1	answer	answer	answer
2	name	localization	command
3	localization	date	date
4	object	number	number
5	date	command	object

Table 2: Top 5 concepts of the semantic annotation for MEDIA, PM-LANG and PM-DOM

Iteration	Dialogues (#utt)	Global WER	Woz side	Callers side
Init	block00 (001-100)	**46.9%**	41.3%	53.2%
0	block01 (101-300)	**15.9%**	7.4%	39.5%
1	block02 (301-500)	**15.8%**	6.9%	37.2%
2	block03 (501-700)	**15.9%**	8.2%	35.6%

Table 3: Word error rate (%) of automatic pre-transcription for each iteration. Woz utterances and callers utterances are distinguished in the last two columns.

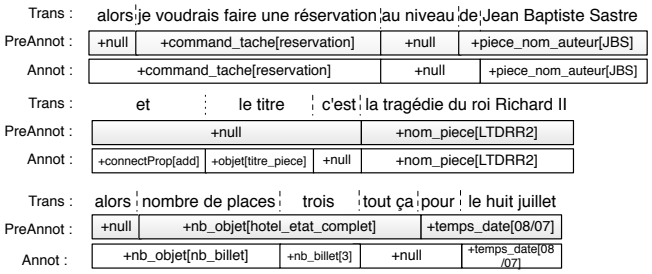


Figure 1: Example of pre-annotation (*PreAnnot*) and its reference annotation (*Annot*) for the given transcription (*Trans*) of a whole user utterance meaning "so I would like to book for Jean Baptiste Sastre and the title it is la tragdie du roi Richard II so number of seats three all that for July eight". For each semantic component, the +/- mode, the name and the value (in square brackets) of the corresponding concept are given.

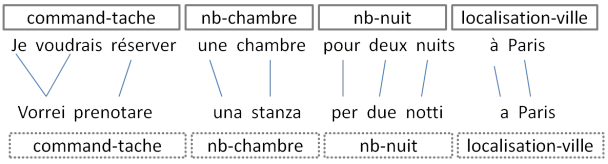


Figure 2: Example of concepts projection from source to target language.

mand_tache concepts are generated by the MEDIA CRF tagger while PM-DOM contains 2459 of them (3758 *answer* concepts out of 3755 to be found, or even 1342 *temps_date* out of 1238 in the reference annotation). However, the drawback of such a direct use of concepts extracted from the source domain is the generation of concepts out of domain with respect to the target data. This is case with concepts very specific of the hotel reservation task (*e.g.* 29 *sejour_nbEnfant*, 2 *chambre_equipement*, 47 *localisation_lieuRelatif_general* etc.) but they represent only 280 semantic pre-annotations (2% of the hypotheses generated). Overall, this first-step system generates 41 different concepts while only 36 of then exists in the PM-DOM semantic specifications with a match of only 23 common concepts for a total of 13,555 automatic semantic annotations.

The second step consists in applying a Named Entity tagger to take advantage of the knowledge coming from the festival program (representing here the general prior knowledge on the domain, which would more often be encompassed into a database). A pretty simple system based on regular expressions is implemented here to limit the time spent in manual developments. The list of searched patterns is composed by the prior-known Named Entities: author names, play titles and locations (305 values). This lead to the automatic generation of 1157 semantic annotations. The final step consists in merging the outputs of the two previous automatic annotations. Rules in the second step are applied so as to have neither overlaps nor splits between concepts. Finally, 14712 concepts are automatically obtained while 17859 are expected (as shown in Table 1).

This pre-annotation system is simplistic but has the advantage to be pretty efficient (low effort/good productivity gain). Obviously new domain-dependent concepts are not exclusively linked to domain Named Entities. A new approach based on Latent Dirichlet Allocation has been proposed in (Camelin et al., 2011) to automatically discover concepts and to provide a semantic annotation without any prior knowledge. However the technique tested on the MEDIA corpus with promising results could not be applied to PORTMEDIA due to time constraints.

Figure 1 depicts an example of a pre-annotation with it reference annotation for a given transcription. It is interesting to note that concept name, value and mode are proposed. This can be helpful for manual annotation and an obvious source of time gain annotation (see Section 3.3.). Also it can be noticed that the automatic pre-annotation system proposes a segmentation in semantic components quite coherent. But this definition of the concept boundaries is a crucial issue for the manual annotation. If too approxima-

Model	Test	Sub	Del	Ins	CER
MEDIA	MEDIA	3.1	15.0	2.3	20.5
	PM-LANG	3.8	13.9	3.1	20.8
PM-LANG	MEDIA	4.7	17.4	3.2	25.3
	PM-LANG	3.6	12.1	3.3	18.9

Table 4: Evaluation (CER %) of Italian models on two different data tests

tive, it can induce a waste of time compare to define them from scratch.

For PM-LANG since we did not have an initial comparable corpus in the target language, we proposed to automatically port the French MEDIA corpus to Italian. Multiple approaches for SLU portability across languages were investigated and compared (Jabaian et al., 2010; Lefevre et al., 2010) and we used the best one to create a new annotated corpus.

The retained approach consists in automatically translating the French MEDIA corpus into Italian and then to port the annotation of the French corpus to the translated one. The automatic translation was carried out by a phrase-based statistical machine translation system trained on a parallel corpus (obtained by manually translating a subset of the French data to Italian[4]). The annotation transfer was based on the automatic alignment between French and Italian sentences. In other words, this method consists in a concept projection using word to word alignments between French and Italian sentences.

Since the French training corpus was already annotated at the chunk level, we proposed to use the alignment information between this corpus and its translation to match directly the semantic concepts to the chunks in the Italian corpus. To do so, we elaborated an algorithm that uses alignment information and boundaries between chunks in French to infer concepts on the Italian side. For each chunk of the French corpus, the algorithm maps the corresponding Italian words based on the alignment information, and then attributes them the corresponding concept. Figure 2 depicts an example of concepts projection using word to word alignment information. This strategy allowed to annotate the entire Italian translated corpus (including both the initial manually-translated part and the subsequent automatically translated subset).

The ported Italian corpus is then used for training a semantic tagger which can be used to perform the first iteration of Italian pre-annotation. For the second and third iterations the manual correction of the pre-annotation is added to the training data. Eventually a brand new model can be trained on the final training data set after the iterative blocks are randomly split into training and test sets.

3.3. Productivity gains

During the transcription and semantic annotation processes we regularly compared the productivity gains due to the semi-automatic transcription/annotation. In this purpose, the following protocol was implemented. Regularly a set of 10 dialogues was transcribed (resp. annotated) twice by two different annotators. The first annotator did the transcription (resp. annotation) from the pre-transcription (resp. pre-annotation) while the second annotator transcribed (resp. annotated) the same 10 dialogues from scratch. The productivity gains are plotted in Figure 3 for the 3 main iterations. The quality of the first iteration for PM-DOM was too bad to be usable by the annotators, so it is considered an initialisation seed instead (and the 0% gain is not represented in the figure).

For the transcriptions, the time to transcribe a whole dialogue was halved using the automatic pre-transcriptions. For semantic annotations, the productivity gains are superior to 50% for Italian and 40% for French.

During the annotation of the Italian data, we evaluated the model used for pre-annotation at each iteration. The influence of the additional manual corrected data can be seen by the decrease of the CER from one iteration to another on the same test set. Otherwise the CER obtained by the same model on the different blocks (represented in Table 5), shows that block01 is more complex than the others (21.9% vs. 20.9% for block00 and 21.1% for block02) which can explain that productivity gain decreases in the second iteration despite the improvement of the pre-annotation model (from 20.8% to 18.7%).

3.4. SLU evaluation on the new test sets

To be comparable to MEDIA, a subset of 200 dialogues was selected from each new corpus as a test set. We evaluate the Italian model used for the pre-annotation and the one trained on the new data. Two test sets are used for this evaluation; the translated MEDIA test set and the PM-LANG test set. For this experiment the evaluation criteria was the concept error rate. Results of this evaluation are given in Table 4.

Results show the robustness of the Italian MEDIA model. Performance of this model on the two test sets are very similar (20.5% vs. 20.8%). This model performs well for a new test set built from scratch as well as for a test set translated from another corpus. The PM-LANG model gives a CER of 18.9% on its corresponding test set, while its performance is lower for the MEDIA test set. This spread in performance may be explained by a difference in concept coverage between the MEDIA and the PM-LANG corpus.

The new domain model trained on the PM-DOM corpus gives a CER of 19.1%, which is very comparable to the performance of the PM-LANG model.

4. High-level semantic annotation

The annotation of the MEDIA corpus with a High Level Semantic (HLS) is investigated: a hierarchical sentence-wise semantic representation. For this purpose, we referred to the MultiModal Interface Language (MMIL) for generating ontology-oriented structures that bears relevant linguistic information from syntax up to discourse (Denis et al., 2010). Therefore, fine-grained features describing dialogue acts, predicates and arguments were defined for the spoken utterances in the corpus.

[4]This small manual translation effort is indeed the weak link of the method and the reason why we consider it to be semi-supervised and not unsupervised.

Iter.	Model	Test	CER
0	MEDIA	PM-LANG	20.8
		block00	20.9
		block01	21.9
		block02	21.1
1	+block00	PM-LANG	18.7
		block01	21.6
2	+block01	PM-LANG	16.4
		block02	19.8

Table 5: CER(%) for each iteration of the semantic annotation process on the data sets (block00 to 02) and on the test set. PM-LANG test set has been compiled afterwards with data from every block.

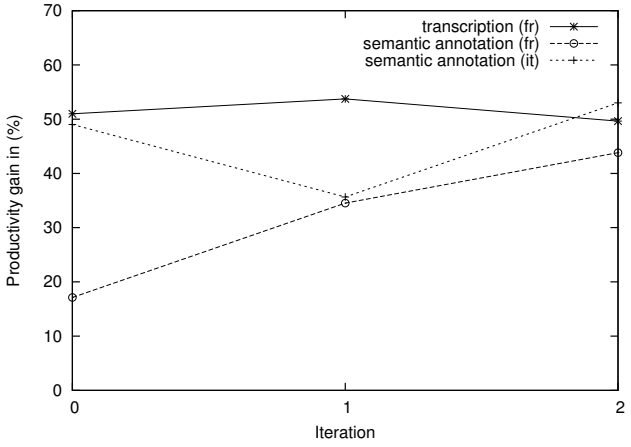
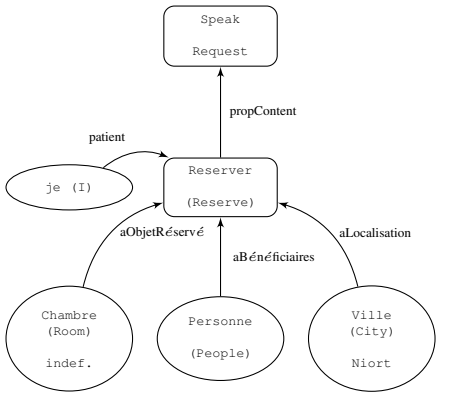


Figure 3: Productivity gains for transcription and semantic annotation in percentage over 3 iterations for sc PM-dom and PM-LANG. block00 of PM-DOM was manually transcribed from scratch and thus is not considered in the plot.

Figure 4 depicts a French utterance represented as a *request to reserve*:

- *The communicative act* is *Request*.
- *The predicate* (or *event*) is *Reserve*
- *The arguments or participants*: the speaker, the beneficiary, the room and the place. The beneficiary and the patient are two different roles, the beneficiary is the person, not necessarily the same speaker, who will use the object reserved (*e.g.* rooms).
- *The features*: Syntactic, semantic (*e.g.* domain specific concepts), pragmatic (*e.g.* referring expressions) and discourse (*e.g.* communicative act).
- *The segmentation*: Each element (*i.e* communicative act, events, participants and features) is mapped to segments of utterances. The table shows how the graph (the HLS representation) is mapped to the utterance and the features of the room are mapped to small segments.



Entities	Segment	Features=Value
Communicative Act:Request	je voudrais ... à Niort	
Main Event:Reserve	faire une réservation	
Participant 1:Pronoun	je	
Participant 2:Chambre	d' une chambre	
	une	refType=indefinite
	chambre	objType=Chambre

Figure 4: HLS representation for the French utterance "je voudrais faire une réservation d' une chambre pour une personne à Niort" (*So I would like to make a reservation for a room for one person in Niort*).

We were specially interested in annotating complex turns, containing overlapped predicates and arguments, as well as elliptical phrases containing either implicit predicates or implicit arguments. First, the annotation guidelines was elaborated[5] and annotated manually a subset of utterances which were supposed to be representative of the most complex aspects of the HLS annotation, in terms of their semantic constituents (Rojas-Barahona et al., 2011).
In a second step, a **baseline** pre-annotation system was elaborated from a rule-based annotation system and we provided metrics to evaluate the automatic annotation (Rojas-Barahona and Quignard, 2011). This baseline is available to be used for comparing the performance of different learning models.
Finally, since the overall objective is the generation of a gold standard, linguist experts are hired to correct the baseline pre-annotation hypothesis. This gold standard will serve for evaluating machine learning methods for this complex semantics.

5. Conclusion

The MEDIA corpus is already available in ELRA catalogue[6]. Two new PORTMEDIA corpora have been collected, transcribed and annotated. They will be added to ELRA catalogue before mid-2012[7], after data validation. The finalization of the feasibility assessment of the proposed high level semantic annotation is still under progress and should be made available with the MEDIA corpus afterwards. So in a very close future the overall MEDIA and

[5]Please refer to http://www.port-media.org/doku.php?id=general_mmil_specifications and http://www.port-media.org/doku.php?id=mmil_for_annotating_media for a full description of the HLS annotation.

[6]http://catalog.elra.info/product_info.php?products_id=998

[7]http://catalog.elra.info/

PORTMEDIA corpora will provide a comprehensive data and tool suite for the evaluation of spoken language understanding systems.

6. Acknowledgements

This work is supported by the ANR funded PORTMEDIA project (ANR 08 CORD 026 01). For more information about the PORTMEDIA project, please visit the project website, www.port-media.org

7. References

T. Anastasakos, J. McDonough, and J. Makhoul. 1997. Speaker adaptive training: A maximum likelihood approach to speaker normalization. In *In: Proceddings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 1043–1046. IEEE.

H. Bonneau-Maynard, S. Rosset, C. Ayache, A. Kuhn, and D. Mostefa. 2005. Semantic annotation of the french media dialog corpus. In *In: Proceedings of Interspeech*, pages 3457–3460, Lisbonne, Portugal, Septembre.

N. Camelin, B. Detienne, S. Huet, D. Quadri, and F. Lefvre. 2011. Concept discovery for language understanding in an information-query dialog system. In *International Conference on Knowledge Discovery and Information Retrieval (KDIR'2011)*, Paris (France).

P. Deléglise, Y. Estève, S. Meignier, and T. Merlin. 2009. Improvements to the LIUM french ASR system based on CMU Sphinx: what helps to significantly reduce the word error rate? In *In: Proceddings of Interspeech*, Brighton (United Kingdom).

A. Denis, L.M Rojas-Barahona, and M. Quignard. 2010. Extending MMIL semantic representation: Experiments in dialogue systems and semantic annotation of corpora. In *In: Proceedings of the Fifth Joint ISO-ACL/SIGSEM Workshop on Interoperable Semantic Annotation (ISA)*, Hong Kong.

L. Devillers, H. Maynard, S. Rosset, P. Paroubek, K. McTait, D. Mostefa, K. Choukri, L. Charnay, C. Bousquet, N. Vigouroux, F. Bechet, L. Romary, J.Y Antoine, J. Villaneau, M. Vergnes, and J. Goulian. 2004. The french media/evalda project: the evaluation of the understanding capability of spoken language dialogue systems. In *In: Proceedings of the Fourth International Conference On Language Resources and Evaluation (LREC)*, Lisbon.

V.V. Digalakis, D. Rtischev, and L.G. Neumeyer. 1995. Speaker adaptation using constrained estimation of gaussian mixtures. *IEEE Transactions on Speech and Audio Processing*, 3(5):357–366.

Y. Estève, T. Bazillon, J.Y. Antoine, F. Béchet, and J. Farinas. 2010. The EPAC corpus: manual and automatic annotations of conversational speech in french broadcast news. In *In: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC)*, Valetta, Malta.

S. Galliano, E. Geoffrois, D. Mostefa, K. Choukri, J. F. Bonastre, and G. Gravier. 2005. The ESTER phase II evaluation campaign for the rich transcription of French broadcast news. In *In: Proceedings of Interspeech*, volume 1, pages 1149–1152, Lisbonne, Portugal, Septembre.

S. Galliano, G. Gravier, and L. Chaubard. 2009. The ester 2 evaluation campaign for the rich transcription of french radio broadcasts. In *In: Proceddings of Interspeech*, Brighton (United Kingdom).

A.L. Gorin, G. Riccardi, and J.H Wright. 1997. How may i help you. *Speech Communication*, 23:113–127.

S. Hahn, M. Dinarelli, C. Raymond, F. Lefèvre, P. Lehnen, R. De Mori, A. Moschitti, H. Ney, and G. Riccardi. 2010. Comparing stochastic approaches to spoken language understanding in multiple languages. *IEEE Transactions on Audio, Speech, and Language Processing*, (99):1–1.

Melvyn J. Hunt. 1990. Figures of merit for assessing connected-word recognisers. *Speech Communication*, 9(4):329–336.

B. Jabaian, L. Besacier, and F. Lefevre. 2010. Investigating multiple approaches for slu portability to a new language. In *In: Proceedings of the 11th Annual Conference of the International Speech Communication Association, Japan, Septembre 2010.*

J. Lafferty, A. McCallum, and F. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *In: Proceedings of the International Conference on Machine Learning*, pages 282–289, Williamstown, MA, USA.

F. Lefevre, F. Mairesse, and S. Young. 2010. Cross-lingual spoken language understanding from unaligned data using discriminative classification models and machine translation. In *In: Proceedings of Interspeech*, Makuhari, Japan.

L. Mangu, E. Brill, and A. Stolcke. 2000. Finding consensus in speech recognition: word error minimization and other applications of confusion networks. *Computer Speech & Language*, 14(4):373–400.

D. Povey and PC Woodland. 2002. Minimum phone error and i-smoothing for improved discriminative training. In *In: Proceddings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages I–105. IEEE.

L. M. Rojas-Barahona and M. Quignard. 2011. An incremental architecture for the semantic annotation of dialogue corpora with high-level structures. a case of study for the media corpus. In *In: Proceedings of the SIGDIAL 2011 Conference*, page 332–334, Portland, Oregon, June. Association for Computational Linguistics, Association for Computational Linguistics.

L. M. Rojas-Barahona, T. Bazillon, M. Quignard, and F. Lefevre. 2011. Using mmil for the high level semantic annotation of the french media dialogue corpus. In *In: Proceedings of the 9th International Conference on Computational Semantics*, Oxford, UK.